International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) e-ISSN: 2395-5325
Volume 4, Issue 1 (January '2018)

A STUDY ON CLUSTER COMPUTING

R.Deepa,
B.Sc (Computer Technology),
Department Of Computer Science,
Sri Krishna Adithya College of Arts and Science,
Coimbatore,Tamilnadu,India.

N. Deepa,
B.Sc (Computer Technology),
Department Of Computer Science,
Sri Krishna Adithya College of Arts and Science,
Coimbatore,Tamilnadu,India.

V. Kowselya,
B.Sc (Computer Technology),
Department Of Computer Science,
Sri Krishna Adithya College of Arts and Science,
Coimbatore,Tamilnadu,India.

Abstract: Cluster computing is defined as the group of linked or loosely coupled computers that are closely working together so that in many respects it forms a single computer. Centralized management approach is followed in the computer clustering. The same task is performed by the nodes in the computer cluster. This is scheduled and controlled by the software. The same operating system and hardware is used in all nodes and different operating system is also used.

Keywords: failover clusters, cluster nodes, bandwidth, seismic, robotics

I. INTRODUCTION

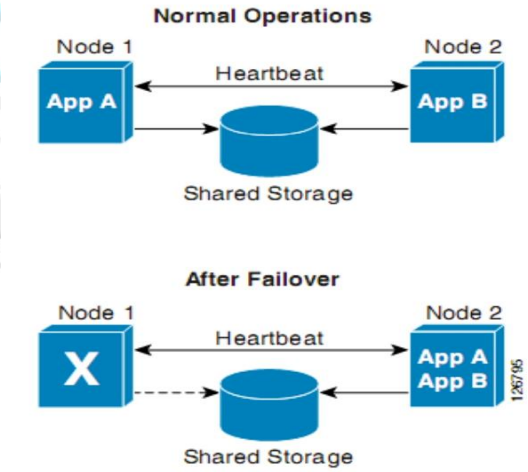
A cluster computing is a type of parallel or distributed processing system, which consists of a collection of interconnected standalone computers cooperatively working together as a single integrated computing resource It supports High Performance Distributed computing (HPDC). Distributed computing techniques are applied in HDPC environment for the solution of computationally intensive applications across network of computers. clustering is the use of multiple computers, storage devices, and redundant interconnections, to form what appears to users as a singly high available system. Computer cluster technology puts cluster of systems together to provide better reliability and performance. To address load balancing, parallel processing, system management, and scalability also be deployed in the clusters. To allow multiple low cost computers to work in a coordinated fashion to process applications for this purpose cluster technologies have been developed.

II.TYPES OF CLUSTER

There are several types of clusters, each with specific design goals and functionality. These clusters range from distributed or parallel clusters for computation intensive applications that are used for protein, seismic, or nuclear modeling to simple loaded- balanced clusters

A. High Availability(HA)or Failover clusters

These High availability clusters are designed for the end-user community to provide uninterrupted of data or services (typically web services).To maintain the availability of services provided by the computer system by replicating servers and services through redundant hardware and software reconfiguration is the main aim of the High Availability(HA) clusters. The purpose of these clusters is to ensures that a single instance of an applications is only ever running on one clusters member at a time but if the cluster member is no longer available , the application will fail over to another cluster member. Nodes can be taken out-of-services for maintenance or repairs. if a node fails, the service can be restored without affecting the availability of the services provided by the clusters(in figure1.1). While the application will still be available ,there will be a performance drop due to the missing node.



(Figure 1.1) Failover clusters

B. Load balancing cluster

Load balancing cluster distributes incoming requests for resources or content among multiple nodes running the same programs or having the same content (figure1.2). The E-commerce and Internet service providers specially uses this type clusters. Who need to resolve differences cargo from multiple input request in real time. Requests are redistributed between the remaining available nodes when a node fails. This type of distribution is typically called web- hosting environment. To increase the reliability, availability, and scalability of application and data resources both high availability and load-balancing cluster technologies can be combined. These are widely deployed for web, mail, news, or FTP services.

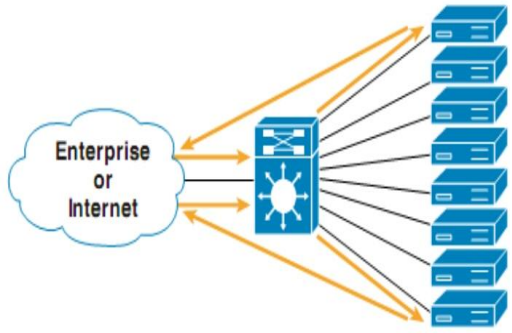


Figure :1.2 (Load balancing cluster)

C. ***Parallel/Distributed processing clusters***

These are systems in which multiple processors share a single memory and bus interface within a single computer. Parallel processing was performed by multiple processors in a specially designed parallel computer. To form a parallel-processing cluster computers are interconnected with advent of high speed, low-latency switching technology. A parallel cluster is a system that uses a number of nodes to simultaneously solve a specific computational or data-mining task. It is specially used for CPU-intensive analytical applications. Beowulf class of cluster is the one of the most common cluster operating systems of clusters. Beowulf cluster is defined as a number of systems whose collective processing capabilities are simultaneously applied to a specific technical, scientific, or business application.

III. COMPONENTS OF CLUSTER

The building blocks of cluster component can be broken down into multiple categories they are the cluster node, cluster operating system, network switching hardware and the node/switch interconnect (in figure3.1) To improve the performance of both the compute node as well as the switching infrastructure significant advances have been accomplished.

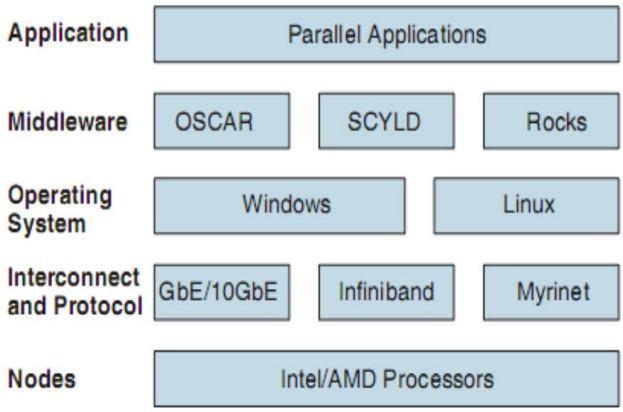


Figure 3.1: Node/switch interconnect

D. ***Application:***

The component application can be called as the input part of the cluster component. These applications run in parallel. In the field of science, This Application of principal component is used to analysis to study of distribution of minor and trace elements in normal human brain. It includes all the various applications that are going on for a particular group. These includes various query running on different nodes of the cluster.

E. ***Middleware:***

For the cluster computing, These are software packages which used to interacts the user with the operating system. These are the building blocks of the cluster computing. In other words we can say that these are the layers of software between applications and operating system. Middleware called software which provides various services required by an application to function correctly. In middleware the software that are used

i. ***OSCAR features:***

- Open Source Cluster Application Resource (OSCAR)
- It is based on Image installation.
- To provide a solution for the challenges of cluster management this project was found.
- It was supported by Red Hat 9.0 and Mandrake 9.0
- The processors which are supported are x86, Itanium(in beta).
- It support in the development of Opteron.
- It is used for High-availability support in alpha testing.
- Interconnects are Ethernet, Myrinet.

ii. SCYLD features:

- It is used for the purpose of commercial distribution.
- The processors which are supported are x86 and opteron.
- It is an single system image design.
- It is called as diskfull and diskless support.

iii. Rocks features:

- Compute node management via Red Hat's kick start mechanism.
- The processors which are supported by Rocks are x86, Opteron, Itanium.
- Ethernet and infiniband is used for the interconnection purpose.

C. Operating system:

There are various operating systems that are supported by cluster computing. Which includes windows, Linux etc. It is an user-friendly interface between the user, the application and cluster hardware. Single system image(SSI) is used in the distributed operating system. The user can interact with the system only when it is single computer. Fault tolerance or error recovery features are implemented. The Distributed operating system for multi-computers is the solaris MC.

D. Interconnect:

10GbE is used for the interconnection between the various nodes for the cluster system.Interconnection like Myrinet etc. With the help of the simple switches we can make interconnections for the small cluster system.

E. Nodes:

The cluster system nodes are the nodes which implies about the different computers that are connected. AMD is a 64-bit processors.

IV. CLUSTER OPERATIONS

Cluster computing operates by various operations such as cluster nodes , cluster network ,cluster applications , message latency and cpu utilization.

A. Cluster nodes:
Cluster computing uses the node technology which has migrated from the conventional tower cases to single rack-unit microproccesors systems and servers that provide a high efficient processor density with a minimal area. As the processor speed and server architecture raised in the performance, it provides the options either in 32-bit or 64-bit processors system. The memory performance, hard disk access speed and storage capacities also increased. The node has two responsibilities: master nodes(or head) and compute nodes(or slave). Master node is responsible for running files as the key system for middleware to route processors and monitor. The compute node in the cluster computing a data storage capability.

B. Cluster network:
Cluster solutions are feasible now-a-days due to the factor such as the high performance servers and the high speed, low latency network switch technology which provides the inter-nodal communication. the cluster computing typically incorporate one or more dedicated switches to communicate between the cluster nodes .with today's low cost per-port for gigabit Ethernet switches, adoption of 10-gigabit Ethernet and the standardization of 10/100/1000 network interfaces on the node hardware,which became a leading interconnect technology for many clusters.

i.network characterization:
There are two primary characteristics in the network operational properties such as bandwidth and delay. The bandwidth is measured in millions of bits per second (mbps) or billions of bits per second (Gbps) .peak Bandwidth is the maximum amount of data that can be transferred in a single unit of time through a single connection. Latency is measured in microseconds or milliseconds and it is the time taken to move a single packet of information in one port and out of another.

ii. Ethernet, fast Ethernet ,gigabit Ethernet and 10-gigabit Ethernet:
Ethernet is the interconnect technology for local area networking. Ethernet as a technology supports speeds varying from 10 mbps to 10 Gbps . It is successfully deployed and operated within the many high-performance cluster computing environments.

c.cluster applications:
cluster computing has a unique and specific applications which are depended upon by the requirements of the computational processes. The different applications such as compute intensive applications, data or i/o intensive applications and transaction intensive applications. Each of these applications has its own set of characteristics and network requirements.

i. compute intensive applications:
It is a application that performs in any computer application that demands a lot of computation cycles. This application is very sensitive to end to end message latency. This is caused by the processors have to wait for instruction messages. The processor have to wait for instruction message to send the resulting data.this application is implemented in the graphic intensive which is the term that implies to any application which demands a lot of computational cycles in which the end result is significant information for the graphical output.

ii. Data or i/o intensive applications:
Data intensive is a term in which it applies to any kind of application that has demands of attached storage facilities. The performance of the data is impacted by the quality of the i/o mechanism. This application is found in the area of data mining , image processing , genome and protein science applications.

iii. Transaction Intensive applications:
Transaction intensive is a term that has a high level of interactive transactions between an application resource and the cluster resources. Many financial banking, human resources and web based applications are came under this category.

d. Message latency:
Message latency is defined as the time taken to send a zero length message from one processor to another processor. It is measured in microseconds. Message latency is made up of aggregate latency incurred at each element within the cluster network and cluster nodes. Network latency is based on the protocol processing latency of message passing interface (MPI) and TCP processes. Cluster nodes are impacted by protocol processing , both TCP/IP processing and the MPI. cluster stability and node synchronization and data sharing are maintained by the message passing technologies such as Parallel Virtual Machine (PVM) OR MPI.

e. CPU Utilization:
The amount of work handled by a CPU or processing rate performed by the computer is termed as the CPU utilization. As the cluster node processes the applications, and protocol processing does not occur, the protocol must interrupt a uniprocessor machine or request a spin lock for a multiprocessor machine. When the request is permitted, CPU cycles are applied to the protocol processes. when more cycles are applied to the protocol processing application processing is suspended.

V.SOFTWARE DEVELOPMENT

The software development in cluster computing includes parallel programming and debugging and monitoring.

i. parallel programming:
The clusters such as web servers use cluster architecture to support a large number of users in which every users request is routed to a specific node, termed as a task parallelism without multi-node. The main goal of the task parallelism is providing rapid user to access the shared data. Automatic parallelization is a parallelization in which the programs continues to remain a technical challenges, but parallel models can be used to effectuate a higher degree of parallelism through the simultaneous execution of separate portion of various processors.

ii.Debugging and monitoring:

- High Performance Debugging Forum (HPDF) which resulted in the HPD specifications is a parallel language primitive or suitable tools implemented for the debugging of parallel programs.
- Application check pointing can be used to restore a system when a node fails during a long multi-node computation. It restores the system to a stable state in which the processing can resume without having to recomputed results.

VI.CLUSTER-AWARE APPLICATIONS

The applications are said to be capable of cluster-aware when it has the following characteristics:

- It uses TCP/IP as a network protocol.
- It maintains data in a configurable location.
- It supports transaction processing.

Some of the cluster-aware applications are: database applications, transaction applications, file and print server applications and groupware applications. The two types of cluster-aware applications are: cluster resource type and not cluster resource type.

i. Clustering applications in Data Mining:

Data mining analysis is a identifying group of related records that can be used as an intiative to explore the further relationship. In this technique it supports the development of population segmentation models such as demographic-based customer segmentation. This technique is actually facilitate the mining process.

ii. Clustering application in text mining:

Text mining also known as text data mining. Text mining is the process of deriving high-quality information from the text. High quality in text mining refers to some combination of relevance, novelty and interestingness. Text mining involves in the process of structuring the input text, deriving patterns and evaluating and interpreting the output. It also includes the text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization and entity relation modeling.

iii. Some other applications of clustering:

The other important applications of clustering are:

- Pattern recognition
- Image analysis
- Bio informatics
- Machine learning
- Voice mining
- Image processing
- Web cluster engines
- Weather report analysis.

VII.ADVANTAGES OF CLUSTER COMPUTING:

Advantages of cluster computing includes:

- Fault tolerance allows for the scalability, and in high performance situations , low frequency of maintenance routines, resource consolidation and centralized management.
- Clustering provides the Reduced Single points of failure through Exchange Virtual Server (EVS) failover functionality.
- It has the ability to perform maintenance and upgrades with limited downtime and ability to easily scale up the cluster to a maximum of seven active EVSs.
- Cluster sampling is less expensive and more quick.
- Cluster sampling procedure enables the users to obtain the information from many areas.
- Clustering servers is completely a scalable solution. Thus the user can add resources to the cluster.
- High availability in the cluster computing helps the user to continue the work with the system if anyone of the components fails, thus some others components can take place .
- Single system image makes the users to not worried about the cluster components, the user only needs to manage a single system image.
- With cluster, large number of components are combined to work as a single entity, thus management become easy.
- Cluster samples permits each and every accumulation to have large samples.

VIII.DISADVANTAGES OF CLUSTER COMPUTING:

- Overlapping effect may take place when adequate number of cases from the stand point of increasing the precision of sample is not selected.
- It is difficult to find fault and determine which component has a problem.
- Cluster computing involves merging different or same components thus it finds a difficult to manage for a non-professional persons.
- Cost is high. Since the cluster needs good hardware and a design.
- Cluster needs more servers and hardware to establish good performance, thus monitoring and maintenance is hard.

IX.CLUSTERING LIMITATIONS

Cluster computing has some limitations such as ,It does not provide protection from:

- Shared storage failures
- Network service failures
- Operational errors
- Site disasters.

X. CONCLUSION

The technologies incorporated with cluster computing includes the host protocol stack-processing , interconnect technologies become the demands of current and emerging applications. Due to the flexibility and high performance it has implemented in many applications such as Google search engine, petroleum reservoir simulation, protein explorer, earthquake simulation and image rendering. With the help of the cluster technology it can be possible to build

better robotics. A robotic cluster is a group of individual robots which are able to share their processing resources, therefore the robots can solve different problems by using the processing units of other robots.

XI.REFERENCES

[1]. https://msdn.microsoft.com/en-us/library/aa369082(v=vs.85).aspx
[2]. https://accountlearning.com/cluster-sampling-definition-advantages-disadvantages/
[3]. https://www.tecmint.com/what-is-clustering-and-advantages-disadvantages-of-clustering-in-linux/
[4]. https://www.slideshare.net/mobile/rajamohd2/cluster-computing-56275811
[5]. https://en.m.wikipedia.org/wiki/Computer_cluster
[6]. https://technet.microsoft.com/en-us/library/aa996423(v=exchg.65).aspx
[7]. http://icl.cs.utk.edu/iter-ref
[8]. http://www.csm.ornl.gov/PR/dell_1.html
[9]. http://www.scribd.com/mobile/doc/5450967/CLUSTER-COMPUTING
[10]. https://googleweblight.com/i?u=https://www.safaribooksonline.com/library/view/high-performance-linux/0596005709/
[11]. https://www.slideshare.net/mobile/poojakhatana1/cluster-computing-28886643
[12]. https://googleweblight.com/i?u=https://www.esds.co.in/blog/cluster-computing-definition-and-architecture-of-a-cluster/&grqid=RQZe1wR9&hl=en-IN&geid=1022
[13]. https://googleweblight.com/i?u=https://msdn.microsoft.com/en-us/library/aa372880(v%3Dvs.85).aspx&grqid=4Zc0utGX&hl=en-IN&geid=1022
[14]. http://vr.sdu.edu.cn/~gb/Architecture/courseware/ClusterComputing.
[15]. M.Baker, A.Apon, R.Buyya , H.Jin, "Cluster Computing and Applications", Encyclopedia of Computer Science and Technology,Vol.45,Marcel Dekker,Aug.2006.
[16]. D.Butenhof, Programming With POSIX Threads, Addison-Wesley,2000.
[17]. R.Buyya (ed.), High Performance Cluster Computing: Systems and Architectures, Prentice Hall,2007.